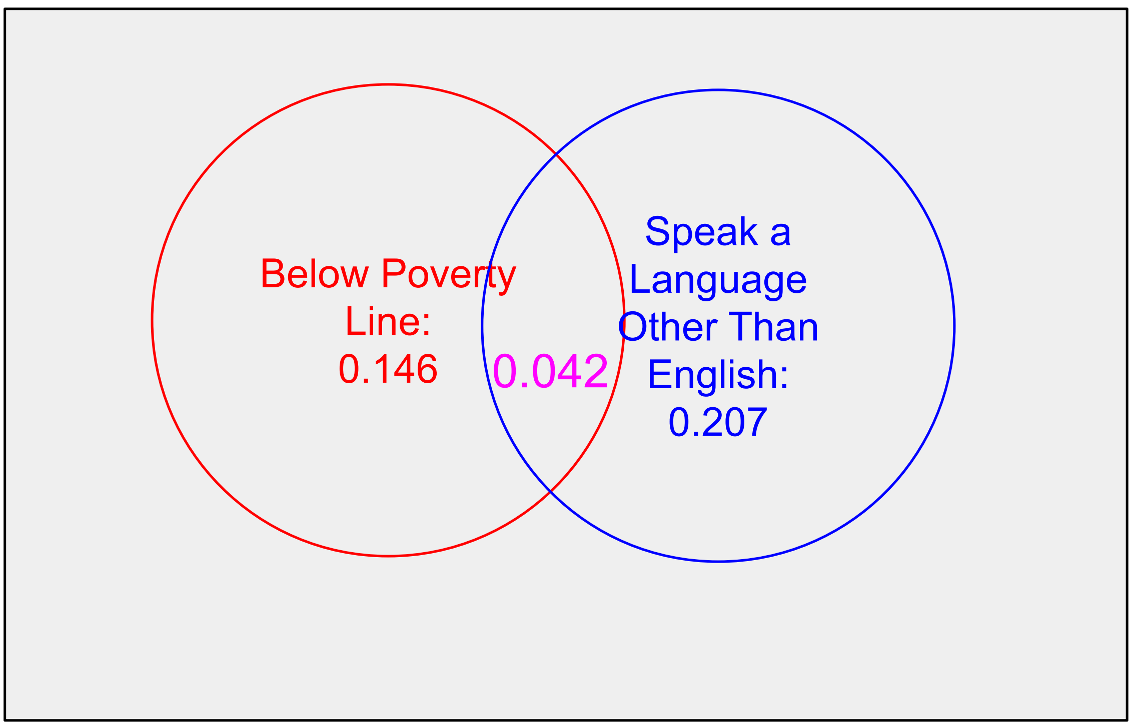
Homework 2

Mark Schulist

1.

a. No, 4.2 of people fall into both categories, so they are not disjoint.

b.



c.

live below poverty line = B

only speak English = E

$$P(B \cap E) = 0.042$$

d.

$$P(B \cup E) = P(B) + P(E) - P(B \cap E) = 0.146 + 0.207 - 0.042 = 0.311$$

e.

$$P(B^c \cap E) = P(B^c)P(E) = 0.854 \cdot 0.207 \approx 0.178$$

f.

$$P(B \cap E) = 0.042$$
$$P(B)P(E) = 0.03022$$

Because these are not equal, they are independent. Living below the poverty line and speaking a foreign language at home are not independent events and knowing one gives us some information about the other.

2.

a. In order to have the first that she gets correct be the fifth question, she needs to get the first 4 incorrect AND get the fifth correct.

$$P(\text{first four incorrect}) = 0.75^4$$
$$P(\text{last one correct}) = 0.25$$
$$P(\text{first four incorrect} \cap \text{last one correct}) = 0.75^4 \cdot 0.25 = 0.0791$$

b. The probability that she gets all of the questions right (assuming independent answers due to randomness) is $P(\text{all correct}) = 0.25^5 = 0.000977$ (not a good idea to guess on a test).

c. The probability of getting at least one correct is the same as not getting all of them wrong.

$$P(\text{at least one correct}) = 1 - 0.75^5 = 0.763$$

3.

a.

$$P(Bl_m \cup Bl_f) = \frac{114}{204} + \frac{108}{204} - \frac{78}{204} = 0.706$$

b.

$$P(Bl_f | Bl_m) = \frac{78}{114} = 0.684$$

c.

$$P(Bl_f | Br_m) = \frac{19}{54} = 0.352$$
$$P(Bl_f | Gr_m) = \frac{11}{36} = 0.306$$

d. They do not appear to be independent as the probabilities change depending on what they condition on. This is especially true for blue-blue pairs, which has a probability much higher than brown-blue and green-blue (where the first color is the male).

4.

Let A = voting for Scott

Let B = having a college degree

$P(A) = 0.53$

$P(A^C) = 0.47$

$P(B|A) = 0.37$

$P(B^C|A) = 0.63$

$P(B|A^C) = 0.44$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

$$P(A|B) = \frac{0.37 \cdot 0.53}{0.37 \cdot 0.53 + 0.44 \cdot 0.47} = 0.487$$

5.

a.

$$\frac{4}{12} \cdot \frac{3}{11} = \frac{1}{11} \approx 0.0909$$

b .

$$\frac{7}{12} \cdot \frac{7}{11} \approx 0.371$$

c.

$$1 - \frac{9}{12} \cdot \frac{9}{11} \approx 0.386$$

d. There are no green socks in the drawer, so it is impossible (zero percent probability).

e.

$$\frac{4}{12} \cdot \frac{3}{11} + \frac{5}{12} \cdot \frac{4}{11} + \frac{3}{12} \cdot \frac{2}{11} \approx 0.288$$

6.

a.

$$\frac{28}{95} \cdot \frac{59}{94} \approx 0.185$$

b.

```
library(tidyverse)
# 1 = hardcover_fiction
# 2 = paperback_fiction
# 3 = hardcover_nonfiction
# 4 = paperback_nonfiction
hf <- rep(1, 13)
pf <- rep(2, 59)
hn <- rep(3, 15)
pn <- rep(4, 8)
books <- c(hf, pf, hn, pn)
iter <- function(books) {
  res <- sample(books, size = 2, replace = F)
  if ((res[1] == 1 || res[1] == 2) && (res[2] == 1 || res[2] == 3)) {
    return(T)
  } else {
```

```r
    return(F)
  }
}

sims <- map(1:10000, ~iter(books)) %>% unlist()
mean(sims)
```

```
[1] 0.2244
```

We are looking for the probability of getting a fiction book first and a hardcover book second. Because these two events have overlapping books (a book can be both a hardcover and fiction), we need to split the problem. We are looking for when the first draw is fiction, so we need to split the problem into hardcover fiction books and paperback fiction books.

Let HCF_1 = getting a hardcover fiction book first

Let PBF_1 = getting a paperback fiction book first

Let HC_2 = getting a hardcover book second

Let

$$P(HCF_1)P(HC_2|HCF_1) + P(PBF_1)P(HC_2|PBF_1)$$

$P(HCF_1) = \frac{13}{95}$ because on this is the first draw

$P(PBF_1) = \frac{59}{95}$ because this is also the first draw

$P(HC_2|HCF_1) = \frac{12}{94}$ as a hardcover book was taken on the first draw, so there is one less HC book and one less book on the shelf overall

$P(HC_2|PBF_1) = \frac{13}{94}$ as a paperback (not a hardcover) book was taken on the first draw, so there are still 13 hardcover books, but one less book overall.

$$\frac{13}{95} \cdot \frac{12}{94} + \frac{59}{95} \cdot \frac{13}{94} \approx 0.224$$

This agrees with the simulation conducted above.

c.

$$\frac{72}{95} \cdot \frac{28}{95} \approx 0.223$$

d. They are very (like, extremely) similar because the sample size is large, so the denominator barely changes (95 to 94). It is the same reason why we assume independence between samples when sampling people in the US, even though they technically are not independent because we don't sample the same person twice.